Final report of OTKA 67783 project

Water is the most important chemical substance. To understand the behavior of this ubiquitous material is essential in science. When a system containing water is large, quantum mechanical description of its properties is impossible. We have to use classical mechanics and electrostatics. This is especially the case when water solvates biopolymers.

The first classical models of water for molecular simulations were created more than 30 years ago. Since then at least 50 models were suggested in the literature [1]. Some of them gained popularity and were applied several thousand times. Such models are the TIP3P [2] which is the water model of biochemical program packages like GROMOS, CHARMM and others. Similarly popular versions are the TIP4P [2] and the SPC/E [3] models. The latter model was applied more than 8000 times in the literature. These models are rigid nonpolarizable molecules. Their parameters are chosen to fit the density, the internal energy and the experimental pair-correlation function of ambient water.

The continuous increase in computing power created a natural urge to replace these models with more realistic ones. The most important property is the polarizability of the model, because the gas phase dipole moment of water, 1.855 D, can become 2.6-8 in condensed phases. Clearly, in the vicinity of a biopolymer the polarizing force varies a lot. Thus, the model must be able to cover the entire range of dipole moments. The first attempt to create polarizable models was published at the end of the 70-s [4]. More reliable models were created in the mid 90-es.

A convenient method was to have three permanent point charges on a fix frame representing the gas-phase molecule and to express the polarization with the addition of a point dipole. This was the idea of the BSV [5] and the DC [6] models. The GCP [7] model applied the same idea but replaced the fixed point charges with Gaussian charges.

An alternative to point dipoles to express polarization is the charge-on-spring method. In this approach one of the charges is connected to the rigid geometry of the model by a classical harmonic spring. The polarizing electrostatic forces pull the spring. Equilibrium is established when these two forces cancel each other. This method was followed by the SWM4 [8,9] model (CHARMM) when, in addition to the three point charges, the charge pair of the Drude oscillator expressed the polarization. Several COS models [10-12] of the GROMOS group used 4 point charges to represent the fix charges and the spring charge connected to the negative charge.

In our previous OTKA project we studied the ice polymorphs of water. It turned out that while the proliferating high pressure structures provide a sophisticated test of potential models, the rudimentary character of the interaction models cannot exploit this possibility yet. So, we turned to the other end of the phase diagram.

We carried out extensive calculations to determine the gas clusters for 13 different water models frequently used in the literature. The major conclusion of this study was that if the negative charge is positioned on the oxygen atom, the molecule will have erroneous structures in gas phase (TIP3P, SPC/E). These models have very distorted dimers and, contrary to quantum chemical calculations and experimental data, flat trimers and tetramers (all atoms are in the same plane). Actually, the calculated phase diagram of water by Vega and coworkers [13] also showed that while the TIP4P model gave qualitatively correct results, only the entire

diagram was shifted by ~40K to lower temperatures, the SPC/E model provided very distorted diagram. Certain ice polymorphs were entirely missing from the phase diagram of the SPC/E model.

In the case of liquids it is more difficult to decide about the correctness of a model because rigid models are fitted to the properties of ambient liquid. We decided to identify structural differences of several models beyond the pair-correlation function. We were searching for trimers in liquid phases of models with different bondlengths, bond angles, and most importantly, different positions of the negative charge. We found that the flat structure characteristic of TIP3P and SPC/E trimers occurs more frequently in these liquids than liquids of other models having the negative charge on the main axis of the molecule at certain distance from the oxygen atom. After all, this is not surprising because to obtain correct gas-phase quadrupole moment, the negative charge should be shifted from the oxygen towards the hydrogen atoms for such simple models.

The other conclusion of gas cluster calculation was that tetrahedral geometries provide lots of artificial low-energy structures not present in nature (TIP5P) [14]. If the polarization of the molecule is confined in a plane, contrary to the almost isotropic polarization of water, the structures will also be erroneous (TIP4P-FQ, SPC-FQ) [15]. The best overall result was given by the GCP model.

These results proved that the Gaussian functions are not only more realistic to represent charge distributions, but provide better results than point charge models. However, we found the application of point dipoles, as in the GCP model, less realistic because they represent interaction sites very far from the source of electrostatic forces which is obviously not the case in condensed water. In addition to this, to derive the Ewald summation for Gaussians is much simpler if there is no point dipole present. Therefore we turned to the charge-on-spring method to express polarization. The model can have only 3 charges, instead of 4(GROMOS) or 5 (CHARMM). This is so because we can connect the spring to any point of the rigid model.

Based on the above experience we created a new model. The first application of Gaussians for describing charge distributions in classical models of water was done by Chialvo and Cummings[16] Later based on this approach a new model of water was developed and tested the GCP model [7]. This model contained three fixed Gaussian charges and a polarizable point dipole. For Gaussian models the partial charges are represented by a spherical charge distribution as

$$\rho_{\alpha,i}(\mathbf{r}) = \frac{q_\alpha}{(2\pi\sigma_\alpha^2)^{3/2}} \exp\left(\frac{-\left|\mathbf{r}-\mathbf{r}_{\alpha,i}\right|^2}{2\sigma_\alpha^2}\right) \quad , \qquad (1)$$

where $\rho_{\alpha,j}$ is the charge density of site α on molecule j. The charge of this site is q_α and the width of the distribution is σ_α. The coulombic interaction energy, U_{qq}, between Gaussian charges can be given as

$$U_{qq} = \frac{1}{2}\sum_{i,j\neq i}^{N}\sum_{\alpha,\beta}\frac{q_\alpha q_\beta}{4\pi\varepsilon_0\left|\mathbf{r}_{\alpha,i}-\mathbf{r}_{\beta,j}\right|} erf\left(\frac{\left|\mathbf{r}_{\alpha,i}-\mathbf{r}_{\beta,j}\right|}{\sqrt{2(\sigma_\alpha^2+\sigma_\beta^2)}}\right) \quad . \qquad (2)$$

It can be seen from Eq.(2) that for point charges ($\sigma \to 0$) we recover the usual formula. This means that program codes written for Gaussian charges can involve point charges without difficulty.

An alternative to polarizable point dipoles is the charge-on-spring approach. In this case a point charge is connected to the molecule by a classical harmonic spring and the equilibrium between the polarizing electrostatic force and the spring force establishes the position of the charge. The formulas are the following: $q\mathbf{l} = \mu = \alpha\mathbf{E} = \alpha\mathbf{F}_e / q$ and $-\mathbf{F}_e = \mathbf{F}_s = -k_s\mathbf{l}$, where q is the charge, $\mathbf{l}$ is the elongation vector of the spring, $\boldsymbol{\mu}$ is the induced dipole, α is the polarizability, $\mathbf{E}$ is the electric field, and $\mathbf{F}_e$ is the electrostatic force which is equal in size with the force of the spring. So, $k_s = q^2 / \alpha$.

The short description of our model is the following. The molecule consists of three Gaussian charges. The centers of the positive charges are placed on the hydrogen atoms, while the negative charge is positioned between the hydrogen atoms and the oxygen atom on the main axis of the molecule (base point). The geometry of the molecule follows its gas phase structure. The OH bondlength is 0.9572 Å and the HOH bond-angle is 104.52°.

In fact, in condensed phases the OH bond is longer and the HOH angle is larger. For rigid models the 2-3° increase of the angle causes only negligible effect. The OH bondlength is more important, especially when relative elongation differences are considered. However, for a rigid model the parameterization overwrites the subtleties of different bondlengths in different phases. Therefore we stuck to the convention followed by all the models above.

The negative charge is -1.210 esu and its distance from the oxygen is 0.2667 Å. This arrangement ensures that the dipole moment of the molecule matches the experimental value, 1.855 D, while the eigenvalues of the quadrupole moment tensor, Q_{xx} = -0.325 DÅ, Q_{yy} = 2.660 DÅ, Q_{zz} = -2.335 DÅ provide the minimum least squares fit to the experimental values: Q_{xx} = -0.134 DÅ, Q_{yy} = 2.626 DÅ, Q_{zz} = -2.493 DÅ .

The square roots of the variances of the Gaussians (σ-s) are fitted parameters. They are 0.45 Å for the positive charge and 0.63 Å for the negative charge. There is some freedom in this choice. The larger the σ-s are, the smaller the internal energy becomes if all the other parameters are the same. In this respect, there is an intangible relationship between the fitted dispersion forces, the polarizibility and the width of the Gaussians.

After a number of tentative calculations and taking into account geometrical considerations, we chose the above values for σ-s as reasonable ones. While the sum of the σ-s can be chosen with some liberty, this is not true for the difference of the negative and positive σ-s. The difference of the two half widths determines the characteristic angle of the gas-phase dimer (the angle between the O-O vector and the bisector of the acceptor molecule) which must be a very important structural feature affecting the correct many-body arrangements in liquid phase as well.

The negative charge is bonded by a classical harmonic spring to its nonpolarized position (base point). In polarized state the electrostatic forces establish mechanical equilibrium with the spring force. Assuming the spring charge is massless and follows the dynamics of the nuclei infinitely fast, this equilibrium should be established in each timestep.

As a dispersion function we applied the following form

$$\varphi(r_{ij}) = A\exp(-Br_{ij}) - C / r_{ij}^6 \quad . \qquad (3)$$

where *A*, *B*, and *C* are parameters to be fitted. We found this form a bit more suitable than the obligatory Lennard-Jones version.

The transferability of the model means that it has the ability to accommodate itself to very different environments. In the case of pure water the two extremes of the phase diagram are

the gas phase and high pressure ice polymorphs. We tested it for our model by decreasing the steepness of the repulsive wing of the dispersion potential assuming that this part of the potential has no effect on the structure and energy of gas clusters but will give good estimate for dense ice phases.

However, we realized that this can only work for very high pressure ice VII because in this phase the closest OO distance is considerably shorter than that in hexagonal ice. In the case of moderately dense ice phases like ice III, ice VI or ice VII at lower pressures the closest OO distances are very close to the same distances in water and hexagonal ice. In the compressed polymorphs of ice the number of neighboring molecules in the second shell, or in the case of ice VII even in the first shell, is larger than in hexagonal ice. The accumulating repulsion of these neighbors lowers the density of these phases if we use the repulsion function suitable for hexagonal ice. In the classical picture it is reasonable to say that the molecules contract under the impact of external pressure.

To handle this problem we devised a self-consistent algorithm. We used two simple exponentials with r^{-6} attraction terms to represent the repulsions in high- and in low-pressure phases. The high-pressure exponential was fixed to be correct for the densest ice VII phases. Then we devised a switch function varying between 0 and 1 to partition the two repulsions for each particle. The argument of the switch function for each molecule was a virial-like force term which characterized the instantaneous compressing force originating from all possible potentials of the system. Using only a few fitting parameters we determined the proper form of the low pressure repulsion-attraction function and the switch function. We showed that this algorithm covers the entire phase diagram of water from gas-phase clusters to high-pressure ice phases with excellent accuracy.

Most of the properties of water are accessible by experimental methods. The most notable exception is the field dependent polarizability. In condensed phases we have no experimental access either to the polarizability, or to the distribution of the molecular dipoles, let alone their higher moments. There are theoretical calculations using different techniques of quantum chemistry but depending on the particular approximation applied, they come to quite different conclusions. The consensus is that the polarizability decreases by increasing field. We carried out a detailed study of our model with respect of its polarizability, dielectric behavior and its properties under the impact of external field.

In gas phase the polarizability of water depends on the orientation of the molecule, but the differences between the diagonal elements of the polarizability tensor are small, so we apply the conventional isotropic value of 1.444 Å^3. It would be complicated to differentiate between directions, let alone we have no idea how these directional values will change under the impact of a finite electric field. We choose the following function to describe the field-dependent dipole moment, μ, of the molecule:

$$\boldsymbol{\mu} = \alpha_\infty \mathbf{E} + \frac{\sqrt{\pi}}{2}(\alpha_0 - \alpha_\infty)\frac{erf(sE)}{sE}\mathbf{E} \qquad (4)$$

where α_0 = 1.444 Å^3 is the polarizability at zero field, α_∞ is the polarizability at infinitely strong field, and s is a parameter to be fitted. The actual polarizability is the following:

$$\alpha = \alpha_\infty + \frac{\sqrt{\pi}}{2}(\alpha_0 - \alpha_\infty)\frac{erf(sE)}{sE} \quad , \qquad (5)$$

and the polariztion self-energy correction is:

$$U_{self} = \frac{1}{2}\left(\alpha_\infty E^2 + (\alpha_0 - \alpha_\infty)\frac{1-\exp(-s^2E^2)}{s^2} \right) \quad . \quad . \quad (6)$$

For α_∞ our choice is 1.0 Å^3. This round value is very close to α_{eff} close to the constant α of the SWM4 models. The value of *s* should be determined by fitting to the calculated dielectric constant. This parameter determines how rapidly the polarizability decreases.

Although the shape of the $\alpha(E)$ function is not known theoretically, Eq. (4) is plausible at low fields, because $\left(\frac{d\mu}{dE}\right)_{E=0} = \alpha_0$ i.e., the polarizability equals the gas-phase value. The other limit at very high field is a threshold value: $\lim_{E\to\infty}\left(\frac{d\mu}{dE}\right) = \alpha_\infty$. Since the induced dipole moment cannot decrease with increasing field, the actual polarizibility at high field is bounded by two extremes. The first is our approach given by the linear limiting form of Eq. (4). (Earlier studies excluded hyperpolarizibility with positive coefficients.) The second relationship $\boldsymbol{\mu} = \alpha(E)\mathbf{E}$ with $\alpha \approx 1/E$ seems less realistic. The polarization function of Ref. 12 is closer to the latter possibility. The problem with $\mu(E_\infty)$ approaching a constant is that the distribution of dipole moments will be highly asymmetric, the Gaussian curve will be cut-off at a maximum value.

There are four criteria to qualify a model created for molecular simulation:
1. represent the most important gas phase properties of a single molecule accurately;
2. use a simple, but physically plausible theory to express the response of the molecule to the impacts of its surrounding;
3. estimate most of the experimental properties of the real material with acceptable accuracy;
4. be as economic and simple as possible in computations.

1.Our model is has correct gas phase dipole moment, and a minimum of least square fit for the quadrupole moment. The charge distributions are not point charges and point dipoles but three Gaussians.
2. The charge-on-spring method worked for many physical problems and it seems that it is reasonably good in this approach as well. The spring connected to a certain point of the geometry automatically ensures the isotropic polarizability of the molecule.
3. The properties of our models are tested on much wider spectrum of properties. Most potentials are tested only for ambient liquid or in some cases for gas-phase dimer properties. We tested our model on the entire phase diagram and obtained better overall mach with experimental data than the rest of the models.
4. We introduced several optimization steps into our algorithm. New integrator, economic estimation of erf functions, etc. With all these optimization our algorithm is slower than the unpolarizable TIP4P only by a factor of 2.6. It is much faster than the SWM4 algorithm and comparable in speed to the COS algorithms. (The latter algorithm performs poorly in certain respect and has obvious errors in its derivation.)

A more detailed account of this project can be find in the listed papers with the exception of the dielectric studies because this will appear in the water and ice issue of PCCP.

At present we do the parallelization of the new code, to increase the speed of the code further. Calculation of the liquid-vapor equilibrium is under way. Results of these calculations will

also be used to determine the spinodal curve and the roughness of the water-vapor surface in cooperation with other researchers.

References

[1] B. Guillot, J. Mol. Liquids., **101**, 219 (2002)
[2] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, J. Chem. Phys., **79**, 926 (1983)
[3] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, J. Phys. Chem., **91** 6269 (1987)
[4] F.J. Vesely, J.Comput. Phys., 24, 361 (1976)
[5] J. Brodholt, M. Sampoli, and R. Vallauri, Mol. Phys., **86**, 149 (1995)
[6] L.X. Dang and T.M. Chang, J. Chem. Phys., **106**, 8149 (1997)
[7] P. Paricaud, M. Psedota, A.A. Chialvo, and P.T. Cummings, J. Chem. Phys.,**122**, 244511 (2005)
[8] G. Lamoureux, A.D. MacKerell, and B. Roux, J. Chem. Phys., **119**, 5185 (2003)
[9] G. Lamoureux, E. Harder, I.V. Vorobyov, B. Roux, and A.D. MacKerell, Chem. Phys. Letter, **418**, 245 (2006)
[10] H.Yu, T. Hansson, and W.F. van Gunsteren, J. Chem.Phys., **118**, 221 (2003)
[11] H. Yu and W.F. van Gunsteren, J. Chem. Phys., **121**, 9549 (2004)
[12] A. P.E. Kunz and W.F. van Gunsteren, J. Phys. Chem, A. **113**, 11570 (2009)
[13] C. Vega, E. Sanz. J.L.F. Abascal and E.G. Noya, J. Cond. Matter, 20, 153101 (2008)
[14] S. W. Rick, J. Chem. Phys., **120**, 6085 (2004)
[15] S.W. Rick, S.J. Stuart, and B.J. Berne, J. Chem. Phys., **101**, 6141 (1994)
[16] A.A Chialvo and P.T. Cummings, Fluid Phase Equilibria, **150-151**, 73 (1998)